# Row-overflow, differential backups, and more

**Paul S Randal**

**Q I recently upgraded an application to run on SQL Server 2005. One of the things I've taken advantage of is the ability to have rows greater than 8,060 bytes so I can allow users to create longer data fields without getting an error from SQL Server. Now that this application is in production, we're having performance issues for some scan queries that used to run fine before the schema change. I've checked the fragmentation of the various indexes and everything is OK. Why are the queries running slowly on SQL Server 2005?**

A The feature you are using, row-overflow, is great for allowing the occasional row to be longer than 8,060 bytes, but it is not well suited for the majority of rows being oversized and can lead to a drop in query performance, as you are experiencing.

The reason for this is that when a row is about to become oversized, one of the variable-length columns in the row is pushed "off-row." This means the column is taken from the row on the data or index page and moved to a text page. In place of the old column value, a pointer is substituted that points to the new location of the column value in the data file.

This is exactly the same mechanism used to store regular LOB (Large Object) columns, such as XML, text, image or varchar(max). Note that if the table schema contains multiple variable-length columns, there is no guarantee that the same column will be pushed off-row when multiple rows become oversized.

This mechanism can create a performance problem. Suddenly a query to retrieve a variable-length column from a single row in a table might need an additional I/O if the column has been pushed off-row (to read in the text page containing the off-row location of the value). If multiple rows are oversized, a query to retrieve the same variable-length column from multiple rows could have unpredictable performance, depending on how many of the values have been pushed off-row.

In your case, a query performing a range scan or table scan for a select list that includes a variable-length column is suffering from poor performance due to row-overflow and its effects. It doesn't matter whether the indexes are perfectly fragmented – when a variable-length column has been pushed off-row, the previously efficient scan is essentially interrupted since a random I/O is necessary to read the text page containing the off-row value.

Row-overflow is still very useful for occasional oversized rows. However, if query performance is critical, it should not be a heavily exploited component in your design.

**Q We've just introduced database mirroring between two failover clusters as a way of getting geo-redundancy for less than the cost of storage area network (SAN) replication. The data centres are within the same city, so we're able to use synchronous mirroring. The problem is that when a failover occurs on the local cluster, the mirrored database fails over to the remote cluster, which is not the behaviour we want. How can we stop this from happening? We want the failover to happen only if the local cluster is unavailable.**

A For increased availability, mirroring is set up with a witness server so failovers occur automatically if the principal becomes unavailable. The idea is that if the entire local cluster goes down, database mirroring will failover to the second cluster and the application can continue.

The problem occurs when a cluster failover happens. The failover takes longer to occur than the default timeout setting of database mirroring. The witness server and mirror server (meaning the active SQL Server instance on the second cluster) agree that they cannot see the principal, and then the mirror server initiates a mirroring failover to the second cluster.

The easiest method for preventing this is to remove the witness server so that database mirroring does not automatically failover if the local cluster goes down. Of course, this reduces availability, as a human is then needed to initiate a failover.

The second option is to alter the default timeout setting of database mirroring. This is the number of once-per-second "pings" that the principal must fail to respond to before it is declared unavailable. This setting is called the partner timeout and has a default value of 10. The current timeout value for the database can be found using the following code:

```
SELECT [mirroring_connection_timeout]
  FROM master.sys.database_mirroring
  WHERE [database_id] = DB_ID ('mydbname');
GO
```

The timeout value can be changed using the following code:

```
ALTER DATABASE mydbname
  SET PARTNER TIMEOUT <timeoutvalue>;
GO
```

For this scenario, the partner timeout needs to be set higher than the usual time it takes a cluster failover to occur on the local cluster. This may be a little tricky to determine given the variability in the time it takes to run recovery on the mirrored database when the cluster failover occurs, but you should be able to determine an upper bound. The problem with this method is that the timeout value may have to be minutes, which may be unacceptable for when a real disaster occurs.

**Q My backup strategy involves full and log backups, but I've heard that I should add differential backups to decrease restore time. I take a full backup once a week and hourly log backups. I tried adding daily differential backups, but one odd thing I've noticed is that the differential backups at the end of the week are close to the same size as the weekly full backup. I was under the impression that they are incremental, just like log backups. Am I missing something?**

A The misunderstanding here is around the nature of differential backups. Unlike log backups, differential backups are not incremental. A differential backup contains all the changed data file extents since the previous full backup (and this applies to database, filegroup, and file level backups).

When an extent (a logical group of eight contiguous data file pages) is changed in any way, it is marked in a special bitmap page called the differential map (or more commonly known as the diff map). There is a diff map for each 4GB chunk of each data file. When a differential backup is taken, the backup subsystem scans all the diff maps and copies all the changed extents, but the diff maps are not reset. This means that if more extents are changed between successive differential backups, the later backups will be larger. The diff maps are only reset when a full backup is performed.

If the application workload is such that the database contents are extensively changed within a short period of time (say, within a week), then a weekly full backup will be almost the same size as a differential backup that was taken just before the next full backup. This explains the behaviour that you're observing.

You are correct in thinking that differential backups offer a way to reduce restore time in a disaster-recovery situation. If the backup strategy is to take weekly full backups and hourly log backups, an up-to-the-minute restore would require the following:

- Take a tail-of-the-log backup (all the logs generated since the most recent log backup).
- Restore the most recent full database backup.
- Restore all log backups, in sequence, since the most recent full database backup.
- Restore the tail-of-the-log backup.

This could require a lot of log backups to be restored, especially if the disaster occurs just before the next full backup is due. (A worst-case scenario would mean 24 + 24 + 24 + 24 + 24 + 24 + 23 log backups to be restored!) By adding daily differential backups to this strategy, the restore sequence changes to this:

- Take a tail-of-the-log backup (all the logs generated since the most recent log backup).
- Restore the most recent full database backup.
- Restore the most recent differential backup.
- Restore all log backups, in sequence, since the most recent differential backup.
- Restore the tail-of-the-log backup.

This could remove the need for restoring a lot of log backups, as restoring a differential backup is essentially the same as restoring all the log backups in the period covered by the differential backup.

The very worst case in a scenario where a daily differential backup is performed would be 23 log backups, even on the last day of the week. The one downside of differential backups not being incremental is that they can take more space, but that's almost always a worthwhile trade-off to reduce restore time.

**Q I have a two-node failover cluster. Each node is running a single instance of SQL Server 2005. I'm following the common advice of setting each instance to only use 50 per cent of the available memory. Now I'm having issues because the workload on both instances needs more memory to maintain the same performance levels. If I remove the memory limitation, or make it higher, I think I'll run into problems when one of the instances fails over and they are both running on just one node. What do you recommend?**

A I'll answer this question for the two-node, two-instance case, but everything below also applies to other multi-instance setups (N-1 failover clusters, where there are N nodes and N-1 SQL Server instances).

Many people experience a high workload (consuming more than 50 per cent of server memory) on both

instances and don't take into account the effect on the workloads when both instances end up running on a single node after a failover occurs. Without any special configuration, it's possible for the memory distribution between the instances to become disproportionate, so one workload runs fine and the other slows to a crawl.

With SQL Server 2000, the recommendation is to limit each instance to a maximum of 50 per cent of cluster node memory. This is because the memory manager in SQL Server 2000 does not respond to memory pressure – if SQL Server takes, say, 80 per cent of the node's memory, it will not give it back. In a failover situation, this means another instance just starting up would only have 20 per cent of the memory available. By limiting both instances to a maximum of 50 per cent of a node's memory, a fail-over instance is guaranteed 50 per cent of the memory. The problem with this is that the workload on each instance is also limited to 50 per cent of the memory.

With SQL Server 2005 (and SQL Server 2008), the memory manager can respond to memory pressure so the 50 per cent maximum is no longer appropriate. But without some kind of limitation, if two instances are running on one cluster node, they may pressure each other until a disproportionate memory distribution is reached.

The answer is to set each instance to have a minimum amount of memory so they cannot be pressured to release too much memory. A common setting for a two-node, two-instance setup is to have each instance configured for a minimum of 40 per cent of the memory. This means that when each instance is running on a separate node, they can consume as much memory as they want. When a failover occurs, each instance is guaranteed a certain amount of memory to preserve a set level of workload performance, with a little left over to be shared between them. Though this means that the performance of both workloads may drop in a failover situation (as expected), they won't be limited at all for the vast majority of the time when each instance is running on a separate cluster node. ■

**Paul S Randal** *is the managing director of SQLskills.com and a SQL Server MVP. He worked on the SQL Server Storage Engine team at Microsoft from 1999 to 2007. Paul wrote DBCC CHECKDB/repair for SQL Server 2005 and was responsible for the Core Storage Engine during SQL Server 2008 development. Paul is an expert on disaster recovery, high availability, and database maintenance and is a regular presenter at conferences around the world. He blogs at SQLskills.com/blogs/paul.*

**Exchange Q & A**

# Outlook Anywhere, the Remote Connectivity Analyzer, and more

**Henrik Walther**

**Q We have just finished deploying Exchange 2007 on Windows Server 2008-based servers in our organisation and things are working very well, with one exception. Even though we have configured Outlook Anywhere (formerly known as RPC over HTTP) following the guidance in the Exchange 2007 documentation on Microsoft TechNet, we can't connect to the Exchange 2007 Client Access servers from an Outlook 2007 client on the Internet, no matter what we try. We have made sure the SAN certificate is trusted by the client and that TCP port 443 is open on the firewall connected to the Client Access servers. Have you ever seen this type of issue?**

A As a matter of fact, I have. You mention that Exchange 2007 was installed on Windows Server 2008-based servers. When a Client Access server has been installed on a Windows Server 2008 server, it's important to keep in mind that Outlook Anywhere won't work properly if IPv6 is enabled on the server. Since IPv6 is enabled by default when Exchange 2007 SP1 is installed on Windows Server 2008, you must make sure to disable it. I've seen several cases where this resolved the issue.

For more information about why Outlook Anywhere and IPv6 on Windows Server 2008 form a bad cocktail, and how you disable IPv6 properly on Windows 2008 servers without breaking Exchange 2007, I recommend you check out the blog post from the Exchange team at Microsoft found at http://www.microsoft.com/uk/exchangeteamblogarchive1. This issue should be fixed with Exchange 2007 SP1 Rollup 4.

**Q** I am currently implementing Outlook Anywhere and Exchange ActiveSync in our Exchange 2007-based messaging environment, and I was wondering if it is somehow possible to test whether Outlook Anywhere will work as expected on the other side of our perimeter network. In addition, I want to make sure the Autodiscover service has been properly configured in our environment. Can you give me any pointers?

**A** Yes, it is possible to test whether Outlook Anywhere is working correctly. Two Microsoft employees (Shawn McGrath from the Exchange Product Group and Brad Hughes from Product Support Services) have created a web-based tool called the Exchange Server Remote Connectivity Analyzer (ExRCA). The tool (in **Figure 1**) should still be considered a prototype, but I have not experienced any bugs or odd behavior whatsoever. The tool can perform Outlook 2007 Autodiscover and RPC/HTTP connectivity tests; it can also test whether Exchange ActiveSync and inbound SMTP mail flow works as expected. Although ExRCA currently isn't supported by Microsoft, I highly recommend it for any remote connectivity tests against Exchange 2007.

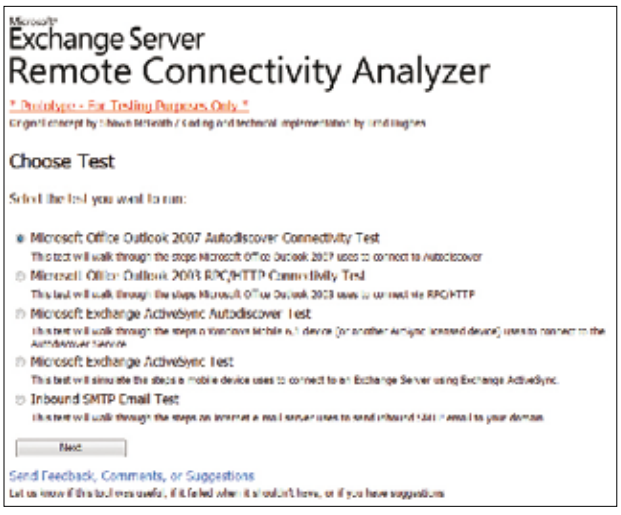


Figure 1 **Exchange Server Remote Connectivity Analyzer start page**

**Q** Our organisation, which uses Exchange Server 2007, is in the planning stages of deploying standby continuous replication (SCR). We want to have a second set of data for each of the mailbox databases created on our non-clustered Exchange 2007 SP1 Mailbox servers in another site. We have been reading a lot about SCR in the Exchange 2007 documentation on Microsoft TechNet but still have a question we haven't managed to get answered there: if we activate an SCR target, will this have the same effect as a Movemailbox with the –ConfigurationOnly parameter specified for all user mailboxes in a particular mailbox database? In other words, only change the Exchange server location in the Active Directory.

**A** Since you're using non-clustered Mailbox servers (otherwise known as a standalone Mailbox server) as source SCR servers, your understanding is correct. Because you will be activating the SCR copy on a different server, database portability will be used. This means that the Exchange server location in Active Directory for the user mailboxes in the respective mailbox database will change. If source SCR servers in your Exchange 2007 environment were either clustered continuous replication (CCR)- or single copy cluster (SCC)-based, and you used a passive node in a failover cluster as the SCR target, you would activate the SCR target with the same name, and the Exchange Server location in Active Directory would not change.

**Q** We have just finalised deployment of Exchange Server 2007 in our enterprise environment and were wondering if it's supported to move the six Exchange 2007 security groups, which were created by Exchange 2007 setup when the forest and domains are prepared for installation of Exchange 2007, to another organisational unit instead of the Microsoft Exchange Security Groups OU, which is created in the root domain.

**A** Unlike Exchange 2000/2003, which didn't allow you to move the Exchange groups to another OU within the forest, Exchange 2007 actually supports doing this. You can see that the six Exchange 2007 security groups (see **Figure 2**) created when the forest is prepared for Exchange 2007 are stamped with two unique properties; the first is a well-known GUID and the second is a distinguished name that can change.

These two properties, and the fact that they are added to the respective forest's OtherWellKnownObjects attribute when setup is run, ensure that

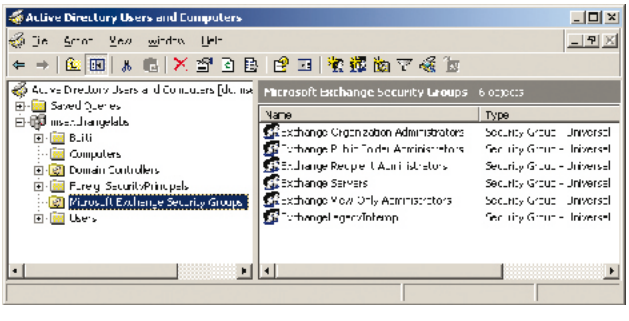


Figure 2 **Exchange Server 2007 security groups**

Exchange will be able to find the security groups anywhere in the forest. So you can go ahead and move the groups anywhere you want to, even to another domain in the forest! Additional details can be found in Ross Smith's excellent Exchange 2007 Permissions FAQ (http://technet.microsoft.com/bb310792) included within the Exchange 2007 documentation on Microsoft TechNet.

**Q Because of some restructuring in our Exchange 2007-based messaging environment, we want to move the file share witness for each of our Exchange 2007 CCR Mailbox servers to another Hub Transport server. Can you provide some guidance on how this is accomplished in a supported fashion?**

A Moving the file share witness from one Exchange 2007 Hub Transport server to another is very straightforward. You simply use the steps that you followed when you initially configured the file share witness for your clustered Mailbox servers. The only difference is the path that you specify to the server. The appropriate steps can be found in the How to Configure the File Share Witness section in the Exchange 2007 documentation on Microsoft TechNet (see http://technet.microsoft.com/bb124922).

By the way, you should know that if you made use of a CNAME record to point to your Hub Transport server when you configured the file share witness, the task would then simply be a matter of you changing the fully qualified domain name (FQDN) of the target host to which the alias in the respective CNAME record points (see **Figure 3**).

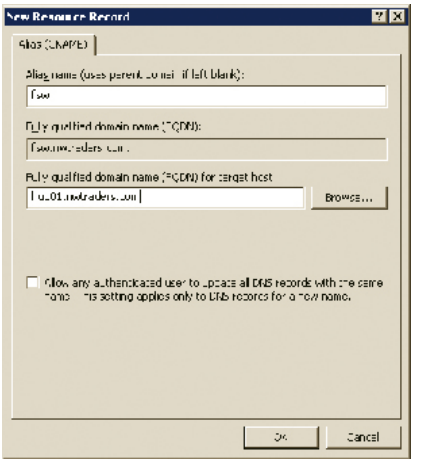


Figure 3 **CNAME record pointing to a target host for a file share witness**

Bear in mind, though, that if you have cluster nodes located in different sites, site resilience guidance from the Exchange Product group has changed (see http://www.microsoft.com/uk/exchangeteamblogarchive2). Basically, the Exchange product group no longer recommends that you use CNAME records in Exchange 2007 Geo-Cluster environments.

**Q We're planning to improve the security settings for the Exchange 2007 messaging servers in our organisation. Part of our security optimisation plan is to encrypt the volumes on which the Exchange databases are located. We wondered whether it is recommended or even supported to store Exchange database files on a volume that has been encrypted using Encrypting File System (EFS) encryption.**

A The answer is a clear no. Placing Exchange 2007 databases on an EFS-based encrypted volume is not supported by Microsoft. In fact, it is unsupported for .edb, .log, .stm (Exchange 2000/2003), .dat, .eml, and .chk files. The primary reason is that this type of encryption results in additional overhead, which significantly affects performance.

To help secure your Exchange 2007 data files further, you should prevent unauthorised access to the Exchange computer and use the S/MIME message format to encrypt message data. Also, if you install Exchange 2007 on Windows Server 2008, consider using BitLocker to protect the volumes.

**Q I've just installed Exchange 2007 SP1 on a Windows Server 2008 server that is also a domain controller. Since I don't use IPv6 in this environment, I disabled it under Network Connections after Exchange 2007 SP1 had been installed, and then I rebooted the server. When it came back online, the Exchange 2007 services no longer started. Error 214, logged in the Application log, contains the following information:**

```
Process MSEXCHANGEADTOPOLOGYSERVICE.EXE
(PID=1712). Topology discovery failed, error
0x80040a02 (DSC_E_NO_SUITABLE_CDC).
```

A I've seen several reports on this behavior. Although it's not good practice to install any of the Exchange 2007 server roles on a Windows Server 2008 server that's also acting as a domain controller, having one or more Exchange 2007 server roles running on a domain controller with IPv6 disabled should work, especially since this is a common scenario in test labs and elsewhere. The solution as of now is to re-enable IPv6 on the server. Rumour has it that Exchange 2007 SP1 Rollup 4 will fix this issue. ■

**Henrik Walther** *is a Microsoft Certified Master: Exchange 2007 and Exchange MVP with more than 14 years of experience in the IT business. He works as a Technology architect for Interprise Consulting (a Microsoft infrastructure Gold partner based in Denmark) and as a technical writer for Biblioso Corporation (a US-based company that specialises in managed documentation and localisation services).*